UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/920,732	08/03/2001	Zbigniew Michalewicz	07100004AA	5103

7590 07/08/2004

McGuire Woods LLP
Suite 1800
1750 Tysons Boulevard
McLean, VA 22102

EXAMINER	
FLEURANTIN, JEAN B	

ART UNIT	PAPER NUMBER
2172	

DATE MAILED: 07/08/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/920,732	MICHALEWICZ ET AL.	
	Examiner	Art Unit	
	Jean B Fleurantin	2172	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *02 April 2004*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-33* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-24,26,28-30,32 and 33* is/are rejected.

7)☒ Claim(s) *25,27 and 31* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Response to Amendment

1. Claims 1-33 remain pending for examination.

Response to Arguments

2. Applicant's arguments filed 2 April 2004 with respect to claims 1-33 have been fully considered but, have been found persuasive only to the extent that the prior art of record does not specifically teach the limitations "building a dictionary based on keywords from an entire text of the documents." However, Riloff discloses such limitations.

In response to applicant's argument on pages 13 and 14, that "the Examiner is using impermissible hindsight reasoning in view of the invention to assert that it would be obvious in view Chundi to cluster documents into groups of clusters where each cluster of the group includes a set of documents containing the same word or phrase. Applicants respectfully submit that the Examiner is using the invention's disclosure to arrive at this assertion. Where in fact, Chundi does not teach or suggest this concept. The use of hindsight in view of the invention is improper," it must be recognized that any judgment on obviousness is in a sense necessarily a reconstruction based upon hindsight reasoning. But so long as it takes into account only knowledge which was within the level of ordinary skill at the time the claimed invention was made, and does not include knowledge gleaned only from the applicant's disclosure, such a reconstruction is proper. See *In re McLaughlin*, 443 F.2d 1392, 170 USPQ 209 (CCPA 1971).

Claim Rejections - 35 USC § 112

3. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claims 1-32 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claim 1, recites the limitation "A method for analyzing and processing documents" in claim. There is insufficient antecedent basis for this limitation in the claim.

Claim 32, recites the limitation "A system for analyzing and processing documents" in claim. There is insufficient antecedent basis for this limitation in the claim.

Claim Rejections - 35 USC § 101

4. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 1-32 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

MPEP 2106 IV.B.2.(b)

A claim that requires one or more acts to be performed defines a process. However, not all processes are statutory under 35 U.S.C. 101. Schrader, 22 F.3d at 296, 30 USPQ2d at 1460. To be statutory, a claimed computer-related process must either: (A) result in a physical

transformation outside the computer for which a practical application in the technological arts is either disclosed in the specification or would have been known to a skilled artisan, or (B) be limited to a practical application within the technological arts.

Claims 1-32, in view of the above cited MPEP section, are not statutory because they merely recite a number of computing steps without producing any tangible result and/or being limited to a practical application within the technological arts. The use of a computer has not been indicated.

Claim Rejections - 35 USC § 103

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1-3, 5-11, 17-24, 26, 28, 29 and 32-33 are rejected under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent No. 6,502,091 issued to Chundi et al. (hereinafter "Chundi") in view of Ellen Riloff et al. "Automated Dictionary Construction for Information Extraction from Text" – 1993 (hereinafter "Riloff").

As per claim 1, Chundi discloses a method for analyzing and processing documents (see col. 4, lines 36-38), comprising the steps of:

analyzing text of the documents for the keywords or a number of occurrences of the keywords and a context in which the keywords appear in the text (see col. 2, lines 8-18, as related terms are grouped into clusters based upon co-occurrence of the terms in the

corpus, terms that co-occur with one or more query terms are used in query expansion and expansion terms are obtained by co-occurrence analysis of all query terms in a query and other terms from top ranked documents); and

clustering documents into groups of clusters based on information obtained in the analyzing step (see col. 2, lines 57-61, as means of applying a clustering algorithm to identify similar query contexts based upon the query keywords to generate context groups that associate keywords with documents accessed by users), wherein each cluster of the groups of clusters includes a set of documents containing a same word or phrase (see col. 2, lines 8-11, as related terms are grouped into clusters based upon co-occurrence of the terms in the corpus, terms that co-occur with one or more query terms are used in query expansion). Chundi does not explicitly disclose steps of building a dictionary based on keywords from an entire text of the documents. However, Riloff discloses automated dictionary construction, (see page 93, col. 1, paragraph 3, line 2 to col. 2, paragraph 1, line 11). It would have been obvious to a person of ordinary skill in the art to modify the combined teachings of Chundi and Riloff with building a dictionary based on keywords from an entire text of the documents. Such modification would have allowed the teachings of Chundi and Riloff to improve the accuracy and the reliability of the system and method for analysis and clustering of documents for search engine.

As per claim 2, Chundi discloses, wherein the clustering step clusters the documents in a catalog tree (see col. 7, lines 51-52 and figure 7, as a set of documents is attached to each node in a multi level context dag).

As per claim 3, Chundi does not explicitly disclose, wherein the clustering step is a static clustering that does not change in response to a user query. The clustering step does not change in response to a user query would have been obvious to one ordinary skill in the art as such would have provided the same result to users making a similar query in order to make the system more consistent.

As per claim 5, Chundi discloses, wherein the analyzing step includes analyzing the documents for statistical information including word occurrences (see col. 6, lines 38-44), identification of relationships between words, elimination of insignificant words and extraction of word semantics (see col. 4, lines 53-61, as a relationships between queries and keywords and documents using the text retrieval system's log, the relationship method identifies all keywords it is related to a set of relevant documents for each context associated with this keyword).

As per claim 6, Chundi discloses, wherein the clustering step is performed recursively, (see col. 36-55).

As per claim 7, the analyzing and clustering steps being performed off line are not explicitly taught by Chundi. However, the analyzing and clustering steps being performed off line would have been obvious to one ordinary skill in the art to have the system of Chundi executed in other to avoid tying up the system when user interaction or processing is contemplated or being performed.

As per claim 8, Chundi discloses the step of generating specific tags for the documents including at least one of document title, document language and summary and the keywords (see col. 3, lines 3-6, as a means of applying a clustering algorithm to identify similar query contexts based upon the query keywords to generate context groups that associate keywords with documents accessed by users).

As per claim 9, Chundi discloses the step of assigning weights to the words and computing the appropriate weights of sentences within the documents (see col. 5, lines 25-26, as various heuristics can be used in order to assign a relevance ranking to a document in the logs).

As per claim 10, Chundi discloses the step of summary generation of the documents, the summary generation being based on the assigned weights to the words and the appropriate weights of the sentences (see col. 5, lines 25-26, as various heuristics can be used in order to assign a relevance ranking to a document in the logs).

As per claim 11, Chundi discloses, wherein the analyzing step is performed on only selected documents which are marked (see col. 1, lines 15-17, as a means of locating documents relevant to a user's information from a collection of documents).

As per claim 17, Chundi discloses, detecting a language of the documents based on frequencies of letter occurrences and co-occurrences in the words (see col. 2, lines 8-18, as related terms are grouped into clusters based upon co-occurrence of the terms in the corpus in

which terms that co-occur with one or more query terms are used in query expansion, and the query expansion terms are obtained by co-occurrence analysis of all query terms in a query and other terms from top ranked documents in the result).

As per claim 18, Chundi discloses, wherein the clustering step is based (i) a best-suited phrase or word from the documents, (see col. 2, lines 8-9).

As per claim 19, Chundi discloses, wherein the analyzing step includes extracting document meta information (see cols. 1-2, lines 67-1, relevance of a document to a given query is computed by the text retrieval system).

As per claim 20, Chundi discloses the steps of generating a cluster hierarchy for the groups of clusters (see col. 5, lines 25-26, as a various heuristic can be used in order to assign a relevance ranking to a document in the log);

generating cluster descriptions, the clustering descriptions including words or phrases that generate a cluster of the groups of clusters and the number of the documents in the cluster (see col. 2, lines 60-61, as a means for applying a clustering algorithm to identify similar query contexts based upon the query keywords to generate context groups that associate keywords with documents accessed by users); and

assigning the documents to elementary clusters and indirect clusters (see col. 7, lines 51-56, a set of documents is attached to each node in a multi-level context dag, the document sets associated with query nodes are the corresponding rel-docs sets from the context discovery steps,

the document set is associated with a general context node is the union of the document sets associated with each of its children, which is similar to the description provided by the applicant (specification on page 41, lines 15-24).

As per claim 21, Chundi discloses, wherein a cluster of the groups of clusters is split into subclusters using statistics to identify best parent cluster and most discriminating significant word in the cluster (see col. 2, lines 58-61, as a means for applying a clustering algorithm to identify similar query context based upon the query keywords to generate context groups that associate keywords with document accessed by users).

As per claim 22, Chundi discloses the step of processing the documents, the processing including: creating reverted index of occurrences of words and phrases in the documents (see cols. 1-2, lines 67-2, as relevance of a document to a given query is computed by the text retrieval system, solely based on the words that appear in the query and the document);

building a directed acyclic graph (see col. 7, lines 38-39, each context group is represented as a directed acyclic graph which is referred to as a "multi-level context directed acyclic graph"); and

extracting a limited number of representative sentences or words or phrases for the document (see col. 5, lines 12-16, as retrieval session contains the query itself and the number of document found to satisfy the query and the list of the documents "ids" or identifiers).

As per claim 23, in addition to claim 1, Chundi further discloses wherein the processing step is independent of the clustering step and is performed in incremental steps, (see col. 5, lines 46-52).

As per claim 24, Chundi discloses, wherein the clustering step includes the steps of creating reverted index of occurrences of words and phrases in the documents (see cols. 1-2, lines 67-2, as relevance of a document to a given query is computed by the text retrieval system, solely based on the words that appear in the query and the document);

building a directed acyclic graph (see col. 7, lines 38-39, each context group is represented as a directed acyclic graph which is referred to as a "multi-level context directed acyclic graph"); and

counting the documents in each group of clusters (see col. 5, lines 18-19, as q_is, are queries and n_i is the number of documents found in the document collection matching q_is).

As per claim 26, Chundi does not explicitly disclose, wherein the analyzing step includes transforming unstructured textual data associated with the documents into structured data in form of tables. However, such is old and well known in the art as exemplified on page 3, lines 11-16 of Applicant's background of the invention section of the instant application. Transforming unstructured textual data associated with documents into a structured data in the form of tables would have been obvious to one ordinary skill in the art in other to provide a structured document with a plurality of associated indexes, therefore providing a faster retrieval time of desired documents.

As per claim 28, Chundi discloses, wherein the documents are divided into different topic domains and restricted to document size (see col. 2, lines 55-58, as a means for partitioning user queries into groups based upon similarity of the query contexts which merging the groups to compute multiple contexts associated with specific query keywords).

As per claim 29, Chundi discloses, wherein a critical size of the documents is determined prior to the analyzing step such that when the critical size exceeds a predetermined size (see col. 2, lines 55-58), the analyzing step only analyzes a first part and a last part of the documents (see col. 1, lines 53-55, as corpus analysis has been used to generate a thesaurus automatically from the content of documents).

As per claim 32, Chundi discloses a system for analyzing and processing documents (see col. 4, lines 36-38), comprising ~~the steps of~~:

a module for analyzing text of the documents for the keywords or a number of occurrences of the keywords and a context in which the keywords appear in the text (see col. 2, lines 8-18, as related terms are grouped into clusters based upon co-occurrence of the terms in the corpus, terms that co-occur with one or more query terms are used in query expansion and expansion terms are obtained by co-occurrence analysis of all query terms in a query and other terms from top ranked documents); and

a module for clustering documents into groups of clusters based on information obtained in the analyzing step (see col. 2, lines 57-61, as means of applying a clustering algorithm to identify similar query contexts based upon the query keywords to generate context groups that

associate keywords with documents accessed by users), wherein each cluster of the groups of clusters includes a set of documents containing a same word or phrase (see col. 2, lines 8-11, as related terms are grouped into clusters based upon co-occurrence of the terms in the corpus, terms that co-occur with one or more query terms are used in query expansion). Chundi does not explicitly disclose a module for building a dictionary based on keywords from an entire text of the documents. However, Riloff discloses automated dictionary construction, (see page 93, col. 1, paragraph 3, line 2 to col. 2, paragraph 1, line 11). It would have been obvious to a person of ordinary skill in the art to modify the combined teachings of Chundi and Riloff with building a dictionary based on keywords from an entire text of the documents. Such modification would have allowed the teachings of Chundi and Riloff to improve the accuracy and the reliability of the system and method for analysis and clustering of documents for search engine.

As per claim 33, Chundi discloses a machine readable medium containing code for analyzing and processing documents (see col. 4, lines 36-38), comprising the steps of:

analyzing text of the documents for the keywords or a number of occurrences of the keywords and a context in which the keywords appear in the text (see col. 2, lines 8-18, as related terms are grouped into clusters based upon co-occurrence of the terms in the corpus, terms that co-occur with one or more query terms are used in query expansion and expansion terms are obtained by co-occurrence analysis of all query terms in a query and other terms from top ranked documents); and

clustering documents into groups of clusters based on information obtained in the

analyzing step (see col. 2, lines 57-61, as means of applying a clustering algorithm to identify similar query contexts based upon the query keywords to generate context groups that associate keywords with documents accessed by users), wherein each cluster of the groups of clusters includes a set of documents containing a same word or phrase (see col. 2, lines 8-11, as related terms are grouped into clusters based upon co-occurrence of the terms in the corpus, terms that co-occur with one or more query terms are used in query expansion). Chundi does not explicitly disclose steps of building a dictionary based on keywords from an entire text of the documents. However, Riloff discloses automated dictionary construction, (see page 93, col. 1, paragraph 3, line 2 to col. 2, paragraph 1, line 11). It would have been obvious to a person of ordinary skill in the art to modify the combined teachings of Chundi and Riloff with building a dictionary based on keywords from an entire text of the documents. Such modification would have allowed the teachings of Chundi and Riloff to improve the accuracy and the reliability of the system and method for analysis and clustering of documents for search engine.

6. Claims 4, 12-16 and 30 are rejected under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent No. 6,502,091 issued to Chundi et al. (hereinafter "Chundi") in view of Ellen Riloff et al. "Automated Dictionary Construction for Information Extraction from Text" – 1993 (hereinafter "Riloff") and further in view of U.S. Patent No. 6,510,406 issued to Marchisio et al. (hereinafter "Marchisio").

The teachings of Chundi are discussed above. As per claim 4, Chundi discloses the step of splitting the groups of clusters into subclusters, the splitting step including finding words which are representative for each of the group of clusters (see col. 2, lines 60-61, as a means for

generating context groups that associate keywords with documents accessed by users); and creating new clusters based on the generating step which corresponds to the top words and a set of phrases (see col. 2, lines 8-18, as related terms are grouped into clusters based upon co-occurrence of the terms in the corpus, terms that co-occur with one or more query terms are used in query expansion and expansion terms are obtained by co-occurrence analysis of all query terms in a query and other terms from top ranked documents). Chundi does not explicitly disclose the steps of generating a matrix containing information about occurrences of the top words in the documents from the groups of clusters. However, Machisio discloses the steps of parsing (breaking) may include processing of tag information associated with html and xml files, and parsing of the electronic information files may further performed, include generating a number of concept identification numbers (concept Ids) corresponding to respective terms (keywords), (see Machisio col. 6, lines 43-52). It would have been obvious to a person of ordinary skill in the art to modify the combined teachings of Chundi, Riloff and Machisio in order to achieve the steps of generating a matrix containing information about occurrences of the top words in the documents from the groups of clusters. Such modification would have allowed the combined teachings of Chundi, Riloff and Machisio to improve the accuracy and the reliability of the system and method for analysis and clustering of documents for search engine, and to provide ability of the search engine to interact with the user and to suggest concepts that may be related to a search, and to browse a list of relevant documents that do not contain the exact terms used in the user query, (see Machisio col. 5, lines 38-42).

As per claim 12, Chundi discloses the claimed subject matter except the claimed feature of wherein the documents are HTML documents. However, Machisio discloses parsing (breaking) may include processing of tag information associated with html and xml files, and parsing of the electronic information files may further performed, include generating a number of concept identification numbers (concept Ids) corresponding to respective terms (keywords), (see Machisio col. 6, lines 43-52). It would have been obvious to a person of ordinary skill in the art to modify the combined teachings of Chundi and Marchisio in order to achieve the steps of wherein the documents are HTML documents. Such modification would have allowed the teachings of Chundi and Machisio to improve the efficiency of the system and method for analysis and clustering of documents for search engine, and to provide ability of the search engine to interact with the user and to suggest concepts that may be related to a search, and to browse a list of relevant documents that do not contain the exact terms used in the user query, (see col. 5, lines 38-42).

As per claim 13, Chundi discloses, wherein the analyzing step includes applying linguistic analysis to the documents, the linguistic analysis being performed on one of titles, headlines and body of the text, and content including at least one of phrases and the words (see col. 1, lines 44-49, word relationships are usually computed by some form of corpus analysis and by using domain knowledge, for example stemming which is a popular method to compute word relationships and reduces words to common roots, as a result documents containing morphological variants of the query words are also included in the response).

As per claim 14, Chundi discloses, wherein the dictionary generates words that describe the contents of the documents, creates indexes for the documents (see col. 1, lines 64-66, as word relationships computed from the top n documents relevant to a query outper-formed the thesaurus generated from the entire corpus), associates the documents with other documents to create concept hierarchy, clusters the documents using a tree-structure of the concept hierarchy and generates a best-suited phrase for cluster description (see col. 2, lines 42-49, a clustering algorithm operative to identify context groups and usage categories which the data mining mechanism is operative to identify query contexts associated with individual queries from the usage logs, partition the queries into context groups having similar contexts, and compute multiple context groups associated with specific query keywords from the usage logs).

As per claim 15, Chundi discloses, wherein the dictionary includes all words appearing in the analyzed documents, and the documents are indexed with the words from the dictionary (see col. 2, lines 4-9, as a means of capturing relationships between keywords by automatic thesaurus construction has previously been studied extensively, some of the earliest work in automatic thesaurus construction was based on term clustering, according to this work, related terms are grouped into clusters based upon co-occurrence of the terms in the corpus).

As per claim 16, Chundi discloses, wherein importance is assigned to each word in the document, the importance being a function of word appearances in the document, position in the document and occurrences in links pointing to the document (see col. 2, lines 8-18, as related terms are grouped into clusters based upon co-occurrence of the terms in the corpus in which

terms that co-occur with one or more query terms are used in query expansion, and the query expansion terms are obtained by co-occurrence analysis of all query terms in a query and other terms from top ranked documents in the result).

As per claim 30, Chundi discloses the claimed subject matter except the claimed feature of wherein the analyzing step includes splitting the documents into separate lexemes including words and hypertext markup language (HTML) tags. However, Machisio discloses parsing (breaking) may include processing of tag information associated with html and xml files, and parsing of the electronic information files may further performed, include generating a number of concept identification numbers (concept Ids) corresponding to respective terms (keywords), (see Machisio col. 6, lines 43-52). It would have been obvious to a person of ordinary skill in the art to modify the combined teachings of Chundi and Marchisio in other to achieve the steps of wherein the analyzing step includes splitting the documents into separate lexemes including words and hypertext markup language (HTML) tags. Such modification would have allowed the teachings of Chundi and Machisio to improve the accuracy and the reliability of the system and method for analysis and clustering of documents for search engine, and to provide ability of the search engine to interact with the user and to suggest concepts that may be related to a search, and to browse a list of relevant documents that do not contain the exact terms used in the user query, (see col. 5, lines 38-42).

Allowable Subject Matter

7. Claims 25, 27 and 31 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

The prior art of record does not teach or suggest in combination with other elements, wherein the clustering step further includes: generating document summaries and statistical data for the groups of clusters;

updating global data by using the document summaries;

generating cluster descriptions of the groups of clusters by finding representative documents in the each cluster of the groups of clusters;

finding elementary clusters associated with the groups of clusters which contain more than a predetermined size of the documents; and

storing the elementary clusters in storage as recited in claim 25.

The prior art of record does not teach or suggest in combination with other elements, wherein the analyzing step includes the steps of: computing a basic weight of a sentence as a

sum of weights of the words in the sentence; normalizing the weight with respect to a length of the sentence; selecting sentences with highest weights;

ordering the sentences with the highest weights in an order which they occur in the input text;

providing a priority to the words by evaluating a measure of particular occurrence of the words in the documents; and

extracting the keywords from the documents which are representative for a given document, the keywords being extracted as follows: for each word s occurring in the document D compute an importance index for s using the formula: Importance

(s , D) = [Priority (s , D) / size (D)] log [N / DF (s)] where N is a number of all the documents and DF(s) is the number of all the documents which contain the word s as recited in claim 27.

The prior art of record does not teach or suggest in combination with other elements, wherein the analyzing step further comprises the steps of:

determining whether there is a next lexeme in the documents;

computing the priorities of all of the words in the documents if the next lexeme is found;

determining which type of information is the lexeme; and

if the documents contain a word lexeme then: obtain an identification of the word from the dictionary;

update statistics of the word occurrence; and

return an ID of the word as recited in claim 31.

Prior Art

8. The prior art of record and not relied on upon is considered pertinent to applicant's disclosure. U.S. Patent No. 5,819,258 issued to Vaithyanathan et al.

U.S. Patent No. 6,446,061 issued to Doerre et al.

U.S. Patent No. 6,480,843 issued to Li

CONTACT INFORMATION

9. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Jean B Fleurantin whose telephone number is 703-308-6718. The examiner can normally be reached on 7:30-6:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John B Breene can be reached on 703-305-9790. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Jean Bolte Fleurantin'

2004-06-18

SHAHID ALAM
PRIMARY EXAMINER